DOCUMENT RESUME

ED 177 186 TM 009 629

TITLE	The Development of a Survey Instrument Codebook.
INSTITUTION	Western Michigan Univ., Kalamazoo. School of Education.
PUB DATE	77
NOTE	13p.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Codification; *Data Processing; *Input Output; Item Analysis; *Questionnaires; Recordkeeping; Records (Forms); *Scoring; *Surveys; Test Construction; Testing; Test Interpretation; Test Results
IDENTIFIERS	*Test Manuals

ABSTRACT

When survey data are analyzed by computer, a codebook is necessary to show how raw scores are translated to computer-usable form. When data are tabulated by hand, the need for a codebook depends upon the number of respondents, the number and complexity of items, and the types of statistical manipulations required. During the process of coding, numerical symbols are assigned to quantify the words, phrases, figures, and symbols recorded as responses to the questionnaire. Coding also describes the location of variables in the computer data file. A carefully compiled codebook should prevent bias in the recording of responses and reduce errors in processing, statistical analysis, or interpretation of results. The codebook contains three sections: (1) introduction--descriptive information about project staff, data storage, and computer specifications; (2) coding classification system--details about recording the alternative responses to an item; and (3) conclusion--explanation of unusual situations, such as the handling of missing data. A codebook is compiled in eight steps: write introduction; edit returned surveys; consult with the researcher; prepare survey instrument; state possible responses; assign codes; list frame numbers; and check for quality control. (A four-page sample codebook is appended.) (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177186

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE DEVELOPMENT OF

A SURVEY INSTRUMENT CODEBOOK

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Center for Educational Research

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Center for Educational Research

©

Western Michigan University 1977
College of Education
Kalamazoo, Michigan

TM009 629

THE DEVELOPMENT OF A SURVEY INSTRUMENT CODEBOOK

INTRODUCTION

All research involves a never-ending fight against error! This battle is not ended when good clean data have been collected. Rather, after the data have been collected, it is necessary to make sure that errors do not creep in as the data are being processed and analyzed.

If data are to be analyzed by computer, it is necessary to develop a codebook. The codebook is like a translation guide which shows how information in the raw or unprocessed form is translated to computer usable form.

When data are to be tabulated or manipulated by hand, the decision as to when one could profit from using a codebook depends upon several factors --

* the number of respondents or sources of data; even when there is a relatively small number of cases, e.g., N=10, the use of coding procedures may avoid or minimize retabulations and frequent handling of the original survey documents.
* the number and type of items in the survey instrument; as the number of items and their degree of complexity increases, coding facilitates calculation of results.
* the type and number of statistical manipulations required; e.g., whether data are needed in raw scores or ranks.

WHAT IS A CODEBOOK?

Coding is the operation of translating raw data into computer-useful format; e.g., it enables quantification of words, phrases, figures and symbols to be assigned numerical symbols which facilitate a variety of processing techniques. It also describes the location of variables in the computer data file.

Thus the type of codebook referred to in this pamphlet deals

primarily with the conversion of raw data from interviews, questionnaires, and similar survey instruments or records to an arrangement which will aid computerized or hand-operated machine analysis and will act as the researcher's guide to the placement of variables within the computer file.

Provided the codebook is compiled correctly, it should prevent bias in the recording of survey instrument responses, eliminate unnecessary errors in the processing and statistical analysis of responses and help researchers avoid faulty interpretation of results.

The coding process can be carried out at any time; simultaneously with the compilation of the survey instrument through to the period just prior to the tabulation of results. The decision as to the best time for coding must be a matter of determining the phase at which one can accomplish the task with the least number of errors and the greatest degree of control.

Researchers who have already field-tested items for the meaning and range of possible answers may decide to have the code physically set out on the final survey instrument; in this way, checking of the answer automatically codes the data. Such pre-coding requires the same procedures for the development of a codebook as coding undertaken after the return of the survey instrument.

PARTS OF A CODEBOOK

Generally codebooks contain three basic sections:

(1) Introduction - this comprises a variety of information essential to the orderly management of data and to the project records. The nine factors included are:

(a) The full name of the project

(b) The name of the researcher and some form of address -- often a phone number

(c) The file name -- a maximum of six characters may be used to indicate the ordered collection of data

(d) The file protection code -- a three digit value which defines who may access, write on or read a file

(e) The project-programmer number -- the number allocated to the project, person or department under which the data file will be stored on the computer

(f) The name of the coder -- the person compiling the codebook should be identified by name and address

(g) Data storage information -- it is advisable to have a written statement describing what will be done with the raw data and other project information once the computer file is compiled

(h) A format statement -- a description of the arrangement of variables in the standard data file

(i) The date of compilation of the codebook

(2) Coding Classification System. Although individual projects may require variations on the theme, one may expect a codebook to have seven feature columns containing

(a) The number of the question or event as it appears on the survey instrument

(b) The variable number as it will appear on the computer

(c) The variable name which may be up to five characters long for use in a Bank Program. The first character must be a letter and all other characters must either be letters or numbers

(d) The description of the variable--derived from the original survey question

(e) The list of possible responses in that variable

(f) The code number to be associated with each possible response

(g) The number of the frame in which the coded response is to be recorded on a mark-sense sheet or computer data record

Occasionally the format statement for each variable is listed but it is advisable to keep the possible responses, the code number and the

frame number/s on the right hand side of the codebook page for easy reference by persons doing the data coding.

(3) Conclusion. At the end of the codebook, sufficient room should be left for notes of explanation or interpretation as to how ambiguities were resolved, e.g., how "missing data" were handled, so that the relevant information remains available to any later data when the data may be required for further examination.

COMPILATION OF A CODEBOOK

The process of compiling a codebook can be described in about eight steps.

(1) Write the codebook introduction.

This section of the codebook is usually compiled with a minimum of difficulty; if problems do arise, they can usually be resolved by consultation with the researcher.

(2) Edit returned surveys.

As complete an understanding of responses as possible is necessary prior to coding. Therefore, the second step is to "edit" the returned survey instruments for accuracy and completeness. Provided this is done early, missing information or errors and ambiguous or hard-to-read responses can be dealt with in a uniform manner.

(3) Consult with the researcher.

Prior to preparing the coding classification system, the coder should discuss the results of the edit with the researcher. Aspects of recording and analysis which warrant consideration include--

(a) the handling of missing data; the choice of a particular method of dealing with missing data depends on the research situation

(b) the levels of measurement; it is not uncommon that insufficient cases appear in some category of a variable to allow for proper statistical analysis. If this has been detected during the editing of returns, the more elaborate code originally considered may be condensed into fewer categories.

(c) special coding requirements; e.g., are any extra variables required -- either subsets or combinations of existing variables? What is to be done when respondents have made multiple checks on an item?

(d) open-ended questions; in the case of free answer or 'open' questions, probes or projective techniques, the conversion to manageable measurement units must be made prior to coding. To determine how the responses will be grouped, a representative sample of responses should be examined, all the responses to the particular question should be copied out and a decision made as to the nature and number of categories to be used.

(4) Prepare the survey instrument.

The first task is that of sorting according to the needs of the project, rejecting instruments which cannot be used, ordering those that can.

The second task, if not previously done at the time of printing the instruments or upon receipt of all the completed questionnaires, is to assign a unique case number to each survey instrument in order from 1 to sample size N. To avoid having "missing data" on the computer file, the number of digits for each assigned number would be the same; i.e., if there are less than 99 cases the first instrument should be assigned the number 01, if there are between 100 and 999 cases the first instrument should be numbered 001.

Individual identification serves several purposes: (a) it allows for detection of missing cases simply be checking the sequence and consecutiveness of identification numbers (b) it permits easy cross-referencing between the original instrument, the coded record and the data file (c) it permits easy checking when new information is to be added to the data file.

(5) State the possible responses.

Responses should be listed in the simplest way, almost invariably in the order in which they appear on the source document as this results

in fewer coder errors, less confusion and increased coder efficiency which will be reflected in the reduction of costs.

(6) Assign codes.

Any code may be assigned to any category of response to an item, but there is usually some efficiency in coding sequentially starting with 0,1,2...N depending on the Number of response categories.

Accuracy and speed in coding is greatly facilitated if, throughout the survey instrument, there is a consistency in the design of codes; e.g., if "no answer" is always "0", "other" is always "9" and so on.

(7) List frame numbers.

The number of frames required for each variable can be readily derived from the number of digits in the variable's codes. The frame list is then used to make up the "format" statement required in the codebook introduction; the format is simply the statement used to specify the desired form of the data to be read by the computer. The format codes are: F for real data not containing a decimal exponent, A for alphameric data and I for integer data. Once the main body of the codebook is compiled, the person handling the statistical manipulations by computer can readily make up the format statement.

(8) Control check for quality.

The final task after designing the codebook is to draw a small sample of completed instruments and try out the coding system.

Code categories must cater to all possible answers on the survey instrument to prevent inconsistencies among coders. Similarly, a final check should be made that the methods of dealing with responses are consistent with the needs of the researcher.

SAMPLE CODEBOOK

The items illustrated in this hypothetical sample have been selected to demonstrate features peculiar to codebooks and are not intended to display internal consistency or orderly development.

PROJECT: Parental attitudes toward career education

RESEARCHER: Director - Dr. Steven Manward Phone: 852-3456

FILE NAME: PATCE

COMPUTER NUMBER: 47003, 30074 PROTECTION: 155

CODER: Mr. Brian Scott Phone: 825-1344

DATA STORAGE: All raw data will be returned to Dr. Manward as soon as the coded data have been placed on the computer and verified as accurate. Raw data to be put on mark-sense sheets for optical scanning.

FORMAT: (I3, I2, 6F1.0, F2.0, 9F1.0)

DATE COMPILED: 11/24/77

SAMPLE CODEBOOK

Item No.	Variable Number	Name	Description	Possible Responses	Code	Frames
-	1	ID	Survey instrument identification	001 002 = to 999	001 002 = to 999	1-3
-	2	School	School attended by the child of this parent	61 62 = to 78	61 62 = to 78	4-5
1.	3	Sexr	Sex of the respondent	No answer Male Female	0 1 2	6
2.	4	Marst	Marital status of respondent	No answer Married Single Divorced Separated Widow/er Other	0 1 2 	7
3.	5	Chnm	Number of children in the family home	Male 0 1 2 = 8 or more Other(NA)	0 1 2 = 8 9	8
-	6	Chnf		Female 0 1 2 = 8 or more Other(NA)	0 1 2 = 8 9	9
4.	7	Resed	Highest level of respondent's education	No answer Elementary Junior high High school College Grad. college Other	0 1 2 3 4 5 9	10

Item No.	Variable Number	Name	Description	Possible Responses	Code	Frames
5.	8	Copin	Opinion of whether all students should have college experience	No answer Strong agree Agree Neutral Disagree Strong disagree Other	0 1 2 3 9	11
6.	9	Exsch	The number of years since the respondent left school ___ years	No answer 1 year 2 years = 12 years 20 years =	00 01 02 = 12 20 =	12, 13
7.	10	Tedis	If teacher of your child were to perform his/her duties inadequately, which of the following do you believe are reasonable methods of handling the situation? (a) dismissal	No answer Yes No Other	0 1 2 9	14
	11	Tenon	(b)non-reappointment	No answer Yes No Other	0 1 2 9	15
	12	Tecs	(c)closer supervision	No answer Yes No Other	0 1 2 9	16
	13	Teco	(d)counseling advice	No answer Yes No Other	0 1 2 9	17
	14	Tetol	(e)tolerate it	No answer Yes No Other	0 1 2 9	18
	15	Teext	(f)extra assignments	No answer Yes No Other	0 1 2 9	19

Item No.	Variable Number	Name	Description	Possible Responses	Code	Frames
	16	Treas	(g)reassignment	No answer	0	20
				Yes	1	
				No	2	
				Other	9	
8.	17	Levca	How much has your level of school influenced your career?	No answer	0	21
				Quite considerably	1	
				Some influence	2	
				No difference	3	
				Other	9	
9.	18	Presj	How long have you worked in your present job?	No answer	0	22
				Less than 2 years	1	
				2-5 years	2	
				6-10 years	3	
				More than 10 years	4	
				Other	9	

NOTES:

(1) Question 2: Researcher requires dichotomous data; discount 0 and 9, leaving categories - married 1
otherwise 2

(2) Question 5: As there were very few persons who marked the strongly agree or strongly disagree categories, it was decided to reduce the Likert scale to a three-point one by combining categories.

ADDITIONAL READINGS

Babbie, E.R. Survey research methods. Belmont, California: Wadsworth Publishing Company, Inc., 1973.

Berdie, D.R. and Anderson, J.F. Questionnaire: Design and use. Metuchen: The Scarecrow Press Inc., 1974.

Galtung, J. Theory and methods of social research. New York: Columbia University Press, 1967.

Goode, W.J. and Hatt, P.K. Methods in social research. New York: McGraw-Hill, 1952.

Selltiz, C.; Jaheda, M.; Deutsch, M. and Cook, S.W. Research methods in social relations. New York: Henry Holt and Company, 1960.